



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/060,579	01/29/2002	D. Matthew Karas	005827.P002	7691

7590 09/02/2005  
Thomas S. Ferrill  
BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN LLP  
Seventh Floor  
12400 Wilshire Boulevard  
Los Angeles, CA 90025-1026

EXAMINER

KERN, MATTHEW C

ART UNIT	PAPER NUMBER
----------	--------------

2654

DATE MAILED: 09/02/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

## Office Action Summary

Application No.

10/060,579

Applicant(s)

KARAS ET AL.

Examiner

Kern Matthew

Art Unit

2654

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

### Status

- 1) ☐ Responsive to communication(s) filed on \_\_\_\_.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

### Disposition of Claims

- 4) ☒ Claim(s) 1-31 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-31 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_ are subject to restriction and/or election requirement.

### Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

### Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some \* c) ☐ None of:
- ☐ Certified copies of the priority documents have been received.
  - ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_.
  - ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).
- \* See the attached detailed Office action for a list of the certified copies not received.

### Attachment(s)

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)  
Paper No(s)/Mail Date 10/31/03;05/06/02.
- 4) ☐ Interview Summary (PTO-413)  
Paper No(s)/Mail Date. \_\_\_\_.
- 5) ☐ Notice of Informal Patent Application (PTO-152)
- 6) ☒ Other: IDS 08/16/02.

## DETAILED ACTION

### ***Claim Rejections - 35 USC § 112***

The following is a quotation of the first paragraph of 35 U.S.C. 112:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

1. Claims 25-31 are rejected under 35 U.S.C. 112, first paragraph, because the specification, while being enabling for the computer readable medium to store software code having one or more attribute filters to detect attributes from an audio information stream, identify the attributes, and assign a time ordered indication with each of the identified attributes, does not reasonably provide enablement for an apparatus. The software engine is purely functional descriptive material. The specification does not enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to use or make the invention commensurate in scope with these claims. This is a single means rejection under Hyatt, 708 F.2d 712, 714-715, 218 USPQ 195, 197 (Fed. Cir. 1983).

### ***Claim Rejections - 35 USC § 102***

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in a patent granted on an application for patent by another filed in the United States before the invention thereof by the applicant for patent, or on an international application

Art Unit: 2654

by another who has fulfilled the requirements of paragraphs (1), (2), and (4) of section 371(c) of this title before the invention thereof by the applicant for patent.

2. Claims 1,5,7-13,15-17,20-21,and 25-28 are rejected under 35 U.S.C. 102(e) as being anticipated by Kanevski et al. (EP 1,076,329).

As per claims 1,5, and 21, Kanevski et al. teach:

- identifying attributes including one or more types of accents (col 4, line 41) and one or more types of human languages (native language, col 4, line 41) from an audio information stream (conversation with a user, figure 5, element 404);
- encoding each identified attribute from the audio information stream into a time ordered index (indicia can be a time stamp, col 11,lines 36-37), each of the identified attributes sharing a common time reference (storing attribute data corresponding to the acoustic feature which is correlated with the at least one user attribute, together with at least one identifying indicia, col 11, lines 30-33); and
- comparing results from different human language models at approximately the same time (French language and American English, col 5, lines 32 and 38, imply being applied at the same time) to generate an integrated time ordered index of the identified attributes (time stamp which correlates the various features to a conversation conducted at a given time, thereby identifying the given transaction, col 11, lines 37-39).

As per claim 7, Kanevski et al. teach instructions which cause the machine to perform further operations comprising:

Art Unit: 2654

- correlating a first identified attribute of the information stream with a second identified attribute having a similar time code (storing attribute data corresponding to the acoustic feature which is correlated with the at least one user attribute, together with at least one identifying indicia, col 11, lines 30-33).

As per claim 8, Kanevski et al. teach wherein the audio information stream comes from an unstructured information source (conversation, col 11, line 17).

As per claim 9, Kanevski et al. teach wherein the audio information stream includes audio-visual data (video information can be included, accompanying audio data, col 16, lines 25-27).

As per claim 10, Kanevski et al. teach where the audio information stream includes speech data (conversation, col 11, line 17).

As per claim 11, Kanevski et al. teach wherein at least one of the identified attributes further comprises a change of accent (different accents which are to be recognized, col 8, lines 52-53, imply detecting a change from one accent to another).

As per claim 12, Kanevski et al. teach wherein at least one of the identified attributes further comprises a change of human language (French language, col 5, line

Art Unit: 2654

32 and American English, col 5, line 38, imply detecting a change between the two languages).

As per claim 13, Kanevski et al. teach wherein at least one of the identified attributes further comprises a discrete spoken word ("pop" vs "soda", col 5, line 43-44).

As per claim 15, Kanevski et al. teach wherein the time ordered index (time stamp, col 11, line 37) includes a start time and a duration in which each identified attribute was conveyed (various features to a conversation conducted at a given time, col 11, lines, implies a start time and duration time for the identified attribute).

As per claim 16, Kanevski et al. teach wherein the common time reference comprises a time indication (time stamp which correlates the various features to a conversation conducted at a given time, thereby identifying the given transaction, col 11, lines 37-39).

As per claim 17, Kanevski et al. teach wherein the common time reference comprises a frame count (25 ms frames with a 10 ms overlap, col 7, lines 31-32).

As per claim 20, Kanevski et al. teach wherein the integrated time ordered index (time stamp which correlates the various features to a conversation conducted at a given time, thereby identifying the given transaction, col 11, lines 37-39) includes data

Art Unit: 2654

from the different human language models (models for each native language, col 5, lines 22-23).

As per claim 25, Kanevski et al. teach an apparatus comprising:

- a software engine having one or more attribute filters to detect attributes from an audio information stream (accent identification, figure 1, element 134), identify the attributes (accent identification, figure 1, element 134), and assign a time ordered indication with each of the identified attributes (storing attribute data (accent) together with at least one identifying indicia (time stamp), col 11, lines 30-34 and lines 37-39, ), the software engine having an index control module to facilitate an integrated time order indexing of the identified attributes (store attribute data, corresponding to acoustic feature, together with identifying indicia, figure 3, element 316); and
- a computer readable medium to store the software engine (computer with appropriate software, col 6, line 38-39).

As per claim 26, Kanevski et al. teach wherein the time ordered index (time stamp, col 11, line 37) includes a start time and a duration in which each identified attribute was conveyed (various features to a conversation conducted at a given time, col 11, lines, implies a start time and duration time for the identified attribute).

As per claim 27, Kanevski et al. teach wherein the one or more attribute filters generate a time ordered index of the audio information stream in real time (performing

Art Unit: 2654

real time storage of an attribute, with an indicia, in the warehouse, is optional, col 17, lines 32-33).

As per claim 28, Kanevski et al. teach wherein the audio information stream passes through the one or more attribute filters a single time (parallel processing set-up of speaker clustering/classification, speaker-independent or class-dependent speech recognition, and accent identification blocks, figure 1, elements 122,128, and 134, respectively).

### ***Claim Rejections - 35 USC § 103***

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

3. Claim 2 is rejected under 35 U.S.C. 103(a) as being unpatentable over Kanevski et al as applied to claim 1 above, and further in view of Bennett et al (US patent application publication 2002/0193991).

As per claim 2, Kanevski et al. does not teach comparing confidence ratings of the different human language models. However, Bennett et al. teach this (recognizer uses algorithms to match what the user says to elements in a speech model, para



[0004] and how confident the recognizer is of each potential match, para [0005]).

Therefore, it would have been obvious to one having ordinary skill in the art at the time of invention to have Kanevski et al. be able to judge the confidence of the interpretation of the audio input of their system so that the user would be sure that when searching the data mining system, all entries for a given language would be retrieved and misclassification could be minimized.

4. Claim 3 is rejected under 35 U.S.C. 103(a) as being unpatentable over Kanevski et al. as applied to claim 1 above, and further in view of Trovato et al (US patent application publication 2002/0163533).

Kanevski et al do not teach generating a transcript including each spoken word. However, the examiner takes Official Notice that it is old and well known in LVSCR systems (speech to text) to generate a real-time transcript of the each word spoken by the user. Therefore, it would have been obvious to one having ordinary skill in the art at the time of invention to have Kanevski et al. be able to generate a transcript of words spoken by the user so that the user could determine whether his speech is being detected by the computer.

Further, Kanevski et al do not teach wherein each spoken word shares the common time reference. However, Trovato et al teach this (timestamp for word I, para [0081]). Therefore, it would have been obvious to one having ordinary skill in the art at the time of invention to have each spoken word in Kanevski et al. share a common time

Art Unit: 2654

reference like in Trovato et al. so that the user could easily locate a keyword or phrase based on timestamp data.

5. Claims 4, 18-19, and 23 are rejected under 35 U.S.C. 103(a) as being unpatentable over Kanevski et al.

As per claim 4, Kanevski et al. do not teach triggering an event to occur upon an identification of unique voice characteristics of a speaker in less than five seconds. However, an artisan would recognize the need to indicate to the user when a unique voice characteristic occurred. Therefore, it would have been obvious to one having ordinary skill in the art at the time of invention to have Kanevski et al. indicate to the user when a unique voice characteristic occurred through an event so that the user could page through many hours of video image/ text data in a short amount of time without worrying whether she is going to miss what she is looking for because she is "scrolling" too fast.

As per claim 18 and 19, Kanevski et al. teach one acoustic feature (MEL ceptstra, col 11, line 25) which is correlated with at least one user attribute (accent, col 11, line 28) with at least one identifying indicia (time stamp, col 11, line 37) conducted at a given time (col 11, lines 37-39). Kanevski et al. do not teach instructions that cause the machine to perform further operations comprising: correlating a first identified attribute (accent) of the information stream with a second identified attribute (language) having a similar time code. However, an artisan would recognize that the spoken

Art Unit: 2654

language of the user could easily be substituted for Mel cepstra in the Kanevski et al. system. Therefore, it would have been obvious to one having ordinary skill in the art at the time of invention to have Kanevski et al. correlate with a time stamp the user's accent with the user's language so that one could determine whether, for example, the language spoken is the mother tongue of the user or a secondary language based on the placement, or lack thereof, of an accent.

As per claim 23, Kanevski et al. do not teach a machine-readable medium that stores instructions, which when executed by a machine, cause the machine to perform operations comprising: converting spoken words in an information stream to written text, the information stream containing audio information. However, the examiner takes Official Notice that it is old and well known in the art to perform this action in large vocabulary continuous speech recognition (LVCSR) systems, also known as speech to text systems. Therefore, it would have been obvious to one having ordinary skill in the art at the time of invention to have Kanevski et al. have speech to text capability so that the user could see a transcript, generated in real-time, of the text that she will be searching.

Further, Kanevski et al. teach generating a separate encoded file for a user attribute, wherein each encoded file shares a common time reference (data warehouse, col 13, line 48, with each warehouse entry understood to mean a file, with an identifying indicia, col 13, line 47, to indicate relative location of one user attribute--accent). Kanevski et al. however, do not teach having an entry/file for each word used in an audio stream. However, the examiner takes Official Notice that an artisan would

Art Unit: 2654

recognize the need to have an entry for each word from the audio stream instead of the user attribute, such as accent, so that the word could be found easily during a query.

Therefore, it would have been obvious to one having ordinary skill in the art at the time of invention to have a file for each word from the audio stream instead of the file for each user attribute taught by Kanevski et al. so that the user of their system would be able to locate individual words, and their associated metadata, quickly.

6. Claims 6 and 22 are rejected under 35 U.S.C. 103(a) as being unpatentable over Kanevski et al. as applied to claim 5 and 21, respectively, above, and further in view of Dharanipragada (6,073,095).

As per claims 6 and 22, Kanevski et al. do not teach instructions which cause the machine to perform further operations comprising: generating a query on one or more of the identified attributes in the time ordered indexed. However, Dharanipragada et al. teach this (retrieve the segments of audio/video that are relevant to the query, col 2, lines 24-25). Therefore, it would have been obvious to one having ordinary skill in the art at the time of invention to have Kanevski et al. have the query ability of Dharanipragada et al. so that specific attributes (accent, language spoken) could be located quickly.

Art Unit: 2654

7. Claim 14 is rejected under 35 U.S.C. 103(a) as being unpatentable over Kanevski et al. as applied to claim 5 above, and further in view of Dremedia (Dremedia—Cutting to the Heart of Technology: Dremedia XML technology).

Kanevski et al. do not teach wherein the identified attributes are encoded via extensible markup language. However, Dremedia teaches automatically tagging an audio transcript with XML tags (automatically attach the appropriate XML tags based on textual output extracted from video and audio streams, para [0003], lines 5-7). Therefore, it would have been obvious to one having ordinary skill in the art at the time of invention to have Kanevski et al.'s system be able to incorporate XML tags so that users could easily link from one part of the transcript to another without having to "scroll" through hours of audio/video data, as taught by Dremedia (para [0003], line 9).

***Allowable Subject Matter***

8. Claims 24, 29, and 31 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

The following is a statement of reasons for the indication of allowable subject matter:

As per claims 24 and 31, none of the prior art teaches a triggering and synchronization module to dynamically trigger a link and synchronize the appearance of the link based upon a transcribed text from the information stream.

As per claim 29, none of the prior art teaches a manipulation module to perform operations on a first set of attributes in order to manipulate a second set of attributes.

### **Conclusion**

9. The prior art made of record and not relied upon is considered pertinent to the applicant's disclosure.

Van Thong et al (6,505,153) teach where text queries match transcript text, which then summons the associated audio content.

Rand et al. (2004/0080528) teach where clicking on a tab button displays links to external web pages.

Kanevski et al. (6,665,644) is the US version of the EP 1,076,329, the main reference used in this office action.

10. Any inquiry concerning this communication should be directed to Mr. Matthew Kern, whose telephone number is (571) 272-7606 or fax number (571) 273-7606. The examiner can normally be reached Mondays-Fridays from 9:30 am to 6 pm.

Art Unit: 2654

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Dr. Talivaldis Smits, can be reached at (571) 272-7628. The facsimile phone number for this Technology Center is (571) 273-8300.

Any inquiry of a general nature of relating to the status of this application should be directed to the Technology Center 2600 receptionist, whose telephone number is (571) 272-2600.

7/21/05

MCK

  
RICHEMOND DORVIL  
SUPERVISORY PATENT EXAMINER